Patterns of Mitochondrial Variation Within and Between African Malaria Vectors, *Anopheles gambiae* and *An. arabiensis*, Suggest Extensive Gene Flow

Nora J. Besansky,*,† Tovi Lehmann,*,† G. Thomas Fahey,* Didier Fontenille,‡ Lawrence E. O. Braack,§ William A. Hawley and Frank H. Collins*,†**

**Centers for Disease Control and Prevention, Division of Parasitic Diseases, Chamblee, Georgia 30341, †Department of Biology, Emory University, Atlanta, Georgia 30322, ‡Laboratoire ORSTOM de Zoologie Medicale, Institut Pasteur, Dakar, Senegal, §Kruger National Park, Skukuza, South Africa and **Kenya Medical Research Institute, Clinical Research Centre, Nairobi, Kenya*

Manuscript received July 10, 1997
Accepted for publication September 15, 1997

ABSTRACT

Anopheles gambiae and *An. arabiensis* are mosquito species responsible for most malaria transmission in sub-Saharan Africa. They are also closely related sibling species that share chromosomal and molecular polymorphisms as a consequence of incomplete lineage sorting or introgressive hybridization. To help resolve these processes, this study examined the partitioning of mtDNA sequence variation within and between species across Africa, from both population genetic and phylogeographic perspectives. Based on partial gene sequences from the cytochrome *b*, *ND1* and *ND5* genes, haplotype diversity was high but sequences were very closely related. Within species, little or no population subdivision was detected, and there was no evidence for isolation by distance. Between species, there were no fixed nucleotide differences, a high proportion of shared polymorphisms, and eight haplotypes in common over distances as great as 6000 km. Only one of 16 shared polymorphisms led to an amino acid difference, and there was no compelling evidence for nonneutral variation. Parsimony networks constructed of haplotypes from both species revealed no correspondence of haplotype with either geography or taxonomy. This trend of low intraspecific genetic divergence is consistent with evidence from allozyme and microsatellite data and is interpreted in terms of both extensive gene flow and recent range expansion from relatively large, stable populations. We argue that retention of ancestral polymorphisms is a plausible but insufficient explanation for low interspecific genetic divergence, and that extensive hybridization is a contributing factor.

Anopheles gambiae and *An. arabiensis* are important components of a malaria-vectorial system encompassing most of sub-Saharan Africa. Of the six closely related Afrotropical sibling species collectively known as the *An. gambiae* complex, these two member species are the most widespread, occurring in sympatry throughout most of sub-Saharan Africa and its offshore islands. In spite of such extensive sympatry, the incidence of naturally occurring interspecies hybrids (recognized chromosomally) is at or below 0.2% (reviewed in COLUZZI *et al.* 1979). The fate of hybrids in nature is unknown, but laboratory experiments have demonstrated their potential as a bridge for gene exchange. Although male F_1 hybrids are sterile, females are fertile and vigorous under laboratory conditions (COLUZZI *et al.* 1979). Interestingly, *An. gambiae* and *An. arabiensis* share cytologically identical paracentric inversions on chromosome 2. If molecular analyses show these inversions to be monophyletic, their presence in both species can be explained either by retention of ancestral polymorphism or by recent gene flow. The gene flow hypothesis is plausible in that chromosomal arrangements from one species have been introgressed into the other by backcrossing in the laboratory (DELLA TORRE *et al.* 1997).

Until quite recently, the only available phylogeny of the *An. gambiae* complex was based on fixed paracentric inversion differences between species (COLUZZI *et al.* 1979; PAPE 1992). In spite of shared autosomal inversions and behavioral similarities, *An. arabiensis* and *An. gambiae* were not predicted to be sister taxa. Their divergent placement on the inversion tree was dictated by species-specific inversions on the *X* chromosome. Recent attempts at phylogenetic reconstruction using DNA sequences have yielded conflicting results. Genetic distance measurements using DNA-DNA hybridization of total single-copy DNA were unsuccessful because differentiation between species was below the limits of resolution of this technique, underscoring their close relationship (N. BESANSKY, G. CACCONE and J. POWELL, unpublished data). However, phylogenetic analyses of sequences from mitochondrial, *X*-linked ribosomal, and two chromosome 2 genes (both outside of the shared inversion) unanimously and strongly supported a sister taxa relationship between *An. gambiae* and *An. arabiensis* (BESANSKY *et al.* 1994; MATHIOPOULOS *et al.* 1995; CACCONE *et al.* 1996), in contradiction

Corresponding author: Nora J. Besansky, Department of Biological Sciences, 317 Galvin Life Sciences Bldg., University of Notre Dame, Notre Dame, IN 46556. E-mail: nora.j.besansky.1@nd.edu

Genetics 147: 1817–1828 (December, 1997)

to the inversion phylogeny. Moreover, mtDNA phylogenies based on the ND4-ND5 genes (BESANSKY *et al.* 1994) and the control region (CACCONE *et al.* 1996) from multiple laboratory strains revealed a paraphyletic relationship between sequences from *An. gambiae* and *An. arabiensis*.

This study attempts to distinguish the current and historical relationships among these taxa, using a comparative population genetic and phylogeographic analysis of mtDNA. It was premised on the expectation that recent or ongoing hybridization leading to interspecies transfer of mtDNA might be distinguished from shared ancestral mtDNA polymorphism, given sufficient intraspecific differentiation over the entire geographic range of each species. The data presented here provide the first detailed picture of the mitochondrial relationships within and among these taxa across Africa, a picture that implicates ongoing hybridization as a contributing factor in their evolutionary dynamics.

MATERIALS AND METHODS

Sampling: Adult *An. gambiae* and *An. arabiensis* were collected from eight locations in Kenya and three locations in Senegal, and additional specimens of *An. arabiensis* were taken from two locations in South Africa that lie outside of the range of *An. gambiae* (COETZEE *et al.* 1993; Figure 1). Morphological identification of anophelines in the field followed GILLIES and DEMEILLON (1968); further identification of species in the *An. gambiae* complex was achieved by a PCR assay (SCOTT *et al.* 1993).

In Kenya, gravid females were collected in May of 1987 by aspiration from the walls or bednets inside dwellings, as previously described (MCLAIN *et al.* 1989). The mosquitoes analyzed were the F_1 progeny, which were reared to adulthood and stored in liquid nitrogen; only one progeny per field-collected female was used.

In Senegal, collections were made from Dielmo and Ndiop between October and December of 1994 by indoor and outdoor night captures on human volunteers as described previously (FONTENILLE *et al.* 1997a,b). Collections at Barkedji were made in September of 1995 by pyrethrum spraying of bedrooms and storerooms (LEMASSON *et al.* 1997). While still in the field, specimens were individually placed into 1.5 ml microcentrifuge tubes with desiccant; long-term laboratory storage was at −20°.

In a section of Kruger National Park, South Africa, at least 9 km from human habitation, a freshwater geothermal spring provides a permanent breeding site for *An. arabiensis* (BRAACK *et al.* 1994). At this site (Malahlapanga; 31°03'E, 22°53'S) in December 1995, mosquitoes seeking a bloodmeal from human volunteers were captured by mouth aspiration. These were sealed in individual gelatin capsules and stored at room temperature with desiccant. Additional *An. arabiensis* specimens from South Africa were obtained from a laboratory strain established from material captured in Mananga (31°50'E, 25°56'S), 350 km south of Malahlapanga.

DNA extraction, amplification, and sequencing: DNA from individual specimens, or legs from specimens, was extracted (COLLINS *et al.* 1987) and resuspended in 50–100 μl water. A 968-bp segment of the mitochondrial NADH dehydrogenase subunit 5 (*ND5*) gene and a 679-bp segment spanning part of the NADH dehydrogenase subunit 1 (*ND1*), the complete $tRNA^{Ser}$ gene, and part of the cytochrome *b* (*cytb*) gene (BEARD *et al.* 1993; GenBank accession # L20934) were PCR amplified in 50-μl reactions containing 1 μl of a 1:10 DNA dilution (or 1 μl undiluted if from legs alone), 50 pmol primers, 5 μl 10× reaction buffer containing 15 mM $MgCl_2$, 200 μM each dNTP (PerkinElmer), and 1.25 U Taq polymerase (Boehringer Mannheim or GibcoBRL). Primers for the *ND5* gene were 19CL (5'-CTTCCACCAATTACTGCTATAACAG-3', positions 6731–6755) and DMP3A (5'-AGGATGAGATGG CTTAGGTT-3', positions 7680–7699). Primers for *ND1-cytb* were DM33C (5'-ACTCTAGCAAGTTTCGAGG, positions 11338–11356) and DM10C (5'-GGTTTAGTCTGGCTAGCT, positions 12000–12017). After 5 min denaturation at 94°, 35 cycles of 15 sec denaturation at 94°, 15 sec annealing at 50°, and 1 min extension at 72° were performed, followed by a 5 min final extension at 72°. PCR products were purified using the Wizard PCR Preps kit (Promega), and cycle-sequenced.

Manual sequencing was performed using the fmol DNA Cycle Sequencing System (Promega) and primers end-labeled with γ-^{33}P. For *ND5*, these primers were the same as those used for PCR amplification. For *ND1-cytb*, the DM33C PCR primer did not perform well for sequencing and was replaced with DM1C (5'-GAGTTCGAGGGACTTTA-3', positions 11751–11767). Because both sequencing primers, DM1C and DM10C, annealed to the same strand, *ND1-cytb* was only sequenced in one direction.

Automated sequencing of *ND5* was performed with the PRISM Dye Deoxy Terminator Cycle Sequencing Kit (Applied Biosystems), and Centricep-purified samples were run on the Applied Biosystems 377 DNA sequencer. Both strands were sequenced, using 6848 (5'-ACTAACCGAAATGAATAACAT ACAG-3', positions 6848–6872) and DMP3A. Sequences have been deposited in GenBank, accession numbers AF020965–AF021023.

mtDNA analysis: The absence of insertions and deletions allowed for unambiguous sequence alignment. Basic sequence statistics were computed with MEGA (KUMAR *et al.* 1993). Haplotype diversity (heterozygosity) was calculated using formula 8.5 of Nei (1987). Estimates of nucleotide variability were not corrected for multiple substitutions because levels of divergence were very low. A program written by T.L. in SAS language (SAS INSTITUTE 1990) was initially used to estimate the average number of pairwise nucleotide differences within and between populations of each species (nucleotide diversity π, equation 10.5 of NEI 1987) as well as the gross and net divergence between them (d_{xy} and d_A of NEI 1987, equations 10.20–21). These and other population genetics parameters were also computed by the program DnaSP 2.0 (ROZAS and ROZAS, 1997). The standard error (SE) of π was the square root of the total (stochastic + sampling) variance per nucleotide site (TAJIMA, 1993). The parameter θ ($= 2N\mu$) was estimated from the number of segregating sites (computed by DnaSP with equation 10.3 of NEI, 1987 and herein called θ_S). Its SE was the square root of the variance per nucleotide site for no recombination (equation 4 of TAJIMA, 1993). Alternatively, the phylogenetic estimate of θ (UPBLUE of FU, 1994 and herein called θ_U) was computed at the Web site of Y.-X. FU, http://hgc.sph.uth.tmc.edu/fu.

To test for neutral mutation, the *D* statistics of TAJIMA (1989) and FU and LI (1993) were computed using DnaSP. TAJIMA'S *D* was calculated using the value of θ_S based on the number of segregating sites. The *D* statistic of FU and LI required an outgroup. This consisted of *ND5* sequences from two field-collected *An. merus*, another sibling in the species complex (GenBank accession numbers AF021024–AF021025).

The extent of nucleotide differentiation between populations was calculated by estimating F_{ST}, where $F_{ST} = 1 - [H_w/H_b]$, and H_w and H_b are the average number of pairwise differences within and between subpopulations, respectively (HUD-

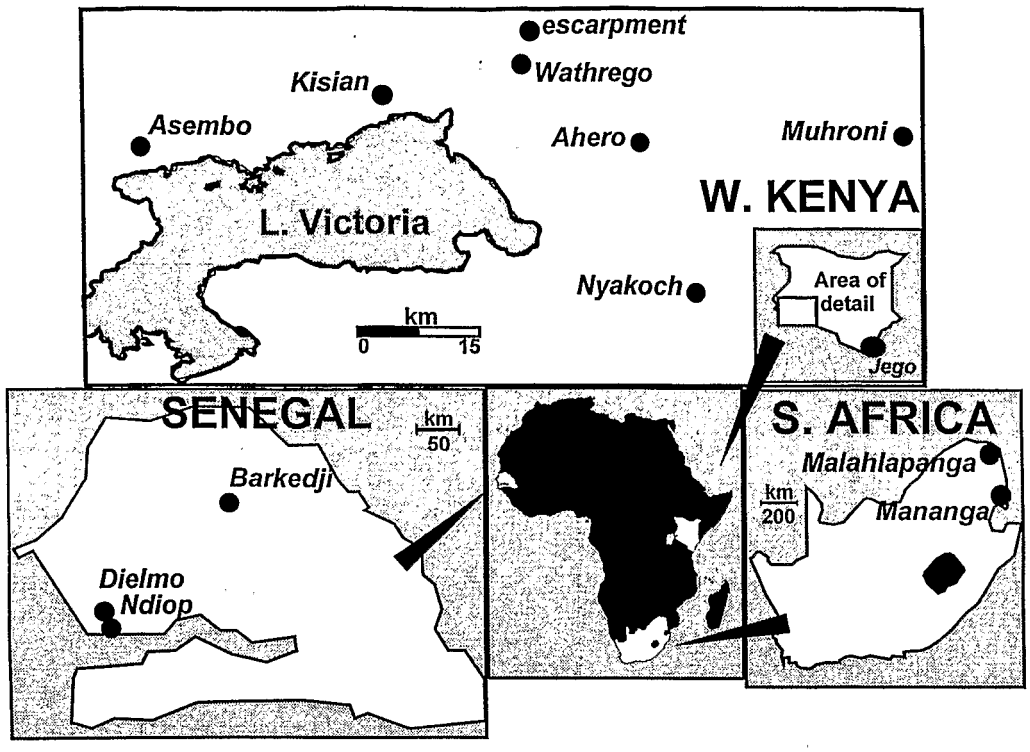

FIGURE 1.—Map of collection sites. Representatives of both species were collected at all sites except those in South Africa (*An. arabiensis* only) and two in Kenya (Ahero, *An. arabiensis* only; Nyakoch, *An. gambiae* only).

SON *et al.* 1992a, Equation 3). To account for differences in sample size, H_w was calculated based on a weighted mean. Significance of F_{ST} estimates was evaluated against the results from 500 random permutations (HUDSON *et al.* 1992b), executed by a program written in SAS by T. LEHMANN. Levels of gene flow (N_m) were estimated by $N_m = H_w/2(H_b - H_w)$ (HUDSON *et al.* 1992a, Equation 4).

Genealogical relationships among mtDNA haplotypes were represented in a hand-drawn network that minimized the number of pairwise mutational differences. The probability of parsimonious (*i.e.*, no unobserved) connections between haplotypes was calculated as in TEMPLETON *et al.* (1992).

Allozyme analysis: The allozyme frequency data reported by MILES (1978) for populations of both species across Africa were analyzed using BIOSYS (SWOFFORD and SELANDER, 1989) as described previously by LEHMANN *et al.* (1996) for *An. gambiae.* Populations of at least 14 mosquitoes from West Africa (The Gambia: Keneba, Mandinari; Senegal: Hanène, Thiès), East Africa (Kenya: Kisumu, Chulaimbo), and Southern Africa (South Africa: Pelindaba; Zimbabwe: Kanyemba) were examined using data from five to six loci. These were the α-naphthyl acetate esterases (*Est-1, Est-3* for both species and *Est-2* for *An. gambiae* only), octanol dehydrogenase (*Odh*), and phosphoglucomutase (Pgm-1, Pgm-2).

RESULTS

Inter- and intraspecific mitochondrial DNA polymorphism: The nucleotide sequence of 665 bp of the *ND5* gene (positions 6896–7560 in the *An. gambiae* reference sequence of BEARD *et al.* 1993) was determined for a total of 65 *An. gambiae* and 56 *An. arabiensis* sampled from three countries: Kenya, Senegal, and South Africa (Figure 1). In addition, the nucleotide sequence of 531 bp containing the $tRNA^{SER}$ gene flanked by portions of the *ND1* and *cytb* genes (positions 11340–11870 in BEARD *et al.* 1993) were determined for the same 37 *An. gambiae* and 23 *An. arabiensis* sampled from Kenya. These two gene regions were chosen because prior restriction site surveys of Kenyan populations from both species identified them as polymorphic (N. BESANSKY, P. MEHAFFEY and F. COLLINS, unpublished data). All sites segregating within *An. gambiae* and *An. arabiensis,* and the sequence variants (haplotypes) defined by them, are shown in Figure 2. All polymorphic sites were either silent codon sites or were conservative replacement sites (position 7290, V → I; 7332, D → N; 7479, V → I; 7485, G → S; 7506, I → V). This, together with the fact that direct sequencing revealed no ambiguities characteristic of heteroplasmy and that no insertion-deletion differences were found within or between species, suggests that the sequenced segments were bona fide mtDNA rather than nuclear-transposed copies.

No sites were fixed at different nucleotides between species. Of the 12 *ND1-cytb* sites that were polymorphic in either species, 11 and three were polymorphic within *An. gambiae* and *An. arabiensis,* respectively. More striking, of 39 *ND5* nucleotide sites polymorphic in either species, 28 and 25 were polymorphic within *An. gambiae* and *An. arabiensis,* respectively. Not only was there substantial shared polymorphism in both species, but the distribution of *ND5* haplotypes by species and geographic origin indicated eight haplotypes common to both species over distances as great as 6000 km (Table 1). The average level of *ND5* sequence divergence between species was only 0.46% per nucleotide site, with a net difference of 0.04% after accounting for within species polymorphism.

Within each species, no single sequence type pre-

		ND5	*ND1-cytb*	
		6666677777777777777777777777777777777	111111111111	
H	F	9999900000111222222233333333444444444445	111111111111	H F
a	r	1126812479023034588901235689002345788 80	455566777777	a r
p	e	3887257559598410528068727016283240935 66	523957236689	p e
#	q	R R R R R	057953993619	# q
a1	(13)	ACCAAACGTTTATTCGATACAAACCTGAAATAAGCCCCT	TGGGAGGAATGT	g-a (1)
*2	(13)	.....A.....C.....	.....G.A.	*-b (15)
g3	(8)	.....C....A.....C.....T.	.....A.	*-c (13)
g4	(1)	.....A.A.....C.....T.	..A...A...A.	g-d (1)
a5	(1)	.....AG.....C.....T.....	C.....A.	*-e (7)
g6	(1)	.....G...AG.....G..C.....	C.....A...A.	g-f (1)
g7	(1)	.....C....A.....C....C.....T.	.....CA.	g-g (3)
a8	(1)	.....A....A.....CA.....T.T.	.....A....A.	g-h (1)
g9	(1)	.....T....C....A.....C.....T.	..AA....G.A.	g-i (1)
a10	(1)	.....A.....T.C.....T.	.....G...	g-j (1)
*11	(7)	.....A.....A.....C.....T.	..A.....A.	g-k (1)
a12	(1)	....G.....A.....T.	CA.....T..A.	g-l (1)
a13	(1)	.....A.....TC.....T.	C.A.....A.	g-m (1)
a14	(1)	.....A.....C.G.....T.	...A.....AA	g-n (1)
a15	(1)	.....A...T.....T.	.....A...A.	g-o (1)
g16	(1)	.....A..G.....C.....G.	C...G.....A.	a-P (11)
a17	(1)	.....AG.....C.....T.T.	*gggaggg*ggg	
a18	(1)	....G.....A.....		
a19	(1)	.....A.....A.....C.....T..		
a20	(1)	.....A.....C.G.....A...T.		
a21	(1)	....G.....A.....C.....G....T.		
a22	(1)	....G.T.....A.....		
a23	(1)	.....C.....		
g24	(1)	TT.....C....A.....C.....T.		
g25	(1)	T.....A.....C..G.....C		
a26	(1)	TT.....A.....C.....A...T.		
g27	(1)	T.....A.....C.....		
g28	(1)	.....A..G.....C.....		
g29	(1)	.....A.....TC.....		
a30	(3)	.....A.....A.		
a31	(2)	.....TA.....		
*32	(5)	.....A.....		
*33	(9)	.....A.....C.....T.		
*34	(2)	.....A.....C..T.....T.		
*35	(2)	.....A.....G...C.....T.		
a36	(1)	.....T.....A.....C.....T.		
a37	(1)	.....A.....T.		
g38	(1)	.....AG.....G..C.....T....		
g39	(1)	.....C....A.....T.C.....T.		
a40	(1)	.....A.....G...C.....T.....		
*41	(4)	.....A.....AG.....C.....T.		
*42	(2)	.....A.....A.....C.....		
a43	(2)	.....C..A.....T.		
a44	(1)	.....AG....G...C.....T.....		
a45	(1)	.....C..A.....		
a46	(1)	.....T.		
g47	(1)	..T.....A.....C.....T...		
g48	(3)	.....AG.....C.....T.		
g49	(1)	.....A.....C...G.....T.		
g50	(1)	..T.....A.....C.....		
g51	(1)	.....A....G.....		
g52	(1)	.....C.....A.....		
g53	(2)	.....C.C....A.....		
g54	(1)	...G...A.....A.....C.....T.		
g55	(1)	.....TA.....C.....T.		
g56	(1)	.....A.....C.....T.T.		
g57	(1)	.....A....G....C..G.....		
g58	(1)	.....A.....C.....A...T.		
g59	(1)	.....A.....CA.....T.		
		ggaga*gaggag*a*agag*g**a*ggaa*g*a*g		

FIGURE 2.—Unique mtDNA haplotypes found for the *ND5* and *ND1-cytb* gene fragments from both species. Only the polymorphic positions are shown, and these are numbered with reference to the published *An. gambiae* mtDNA sequence (BEARD *et al.*, 1993; GenBank accession no. L20934). Dots represent identity with respect to the first sequence listed. Each haplotype was assigned a number that was preceded by the letter a or g if it was unique to *An. arabiensis* or *An. gambiae*, respectively; if shared by both species, it was preceded by an asterisk. The frequency, given in parentheses, indicates the number of times the haplotype was found in the total sample. The five positions in *ND5* in which bp substitutions would result in an amino acid replacement are indicated by an R above the sequence. Below the list of sequences, an asterisk marks positions that are polymorphic in both species; an a or g marks positions that are polymorphic only in *An. arabiensis* or *An. gambiae*, respectively.

TABLE 1

Geographic distribution of *ND5* haplotypes from *An. gambiae* and *An. arabiensis*

	An. gambiae		*An. arabiensis*	
Collection site	Sample size	mtDNA haplotypes[a]	Sample size	mtDNA haplotypes[a]
Kenya				
Asembo	8	***2***, ***11***, ***32***, ***33***, 47, 48 (2), 49	4	1 (2), ***2***, 40
Escarpment	6	***33***, ***42***, 50, 51, 52, 53	2	1, ***41***
Ahero	—	—	4	1, ***11***, ***42***, 43
Kisian	4	***2***, ***11***, ***32***, 54	2	1, 44
Wathrego	6	***2***, ***33*** (3), 48, 53	5	1 (2), ***11***, 43, 45
Nyakoch	4	***2***, ***11***, 55, 56	—	—
Jego	5	***32*** (2), ***41*** (2), 57	2	1 (2)
Muhroni	4	***33*** (2), 58, 59	4	1, ***2***, ***41***, 46
Senegal				
Dielmo	10	***2*** (4), 3 (2), 4, 16, ***35***, 39	9	5, 17, 18, 19, 20, 21, 22, 23, 26
Ndiop	9	***2***, 3 (2), 6, 7, 24, 25, 27, ***34***	9	1 (2), 10, ***11***, 12, 13, 14, 15, 37
Barkedji	9	***2***, 3 (4), 9, 28, 29, 38	8	**1**, ***2***, 8, ***33*** (2), ***34***, ***35***, 36
South Africa				
Malahlapanga	—	—	5	***11***, 30 (3), ***32***
Mananga	—	—	2	31 (2)

Haplotypes shown in bold italic are present in both species.
[a] Values in parentheses are frequency.

dominated. Estimated haplotype diversity was quite high in all populations sampled, although certain haplotypes (*e.g.*, *2* and *11*) were geographically widespread (Table 1). Among the 65 *An. gambiae ND5* sequences, 33 haplotypes were found, 26 of which were singletons (represented in a single individual). Similarly, 11 of 15 *ND1-cytb* haplotypes were singletons. In *An. arabiensis*, the same pattern was detected. Among 54 *ND5* sequences (excluding the two derived from the Mananga colony), 32 haplotypes were found, 25 of which were singletons. The two mosquitoes from the Mananga colony shared the same haplotype, which differed slightly from the 32 sampled from natural populations. For the 23 *An. arabiensis ND1-cytb* sequences, there were no singletons among four haplotypes, yet diversity was still relatively high.

This diversity of haplotypes was achieved with relatively slight differentiation in pairwise comparisons (Table 2). Thus, the average number of pairwise nucleotide differences (π) for *An. gambiae ND5* was 0.38% per site, or 2.5 nucleotides (nt) per sequence. Similarly, π for *An. arabiensis* was 0.46% per site for *ND5*, or 3.1 nt per sequence. In neither species did any pair of haplotypes differ by more than 7 nt.

Tests of neutral mutation: Under the infinite-sites model of neutral mutations and a population at equilibrium with respect to mutation and drift, π and θ should both equal $2N\mu$ for mtDNA, where N is the effective population size and μ is the rate of mutation per site per generation (Nei, 1987). In *An. gambiae* and *An. arabiensis* populations, values of θ_S and θ_U were generally higher than those for π (Table 2). Nevertheless, most values of Tajima's D (Tajima, 1989) and Fu and Li's D (Fu and Li, 1993) were consistent with neutral mutation of *ND5* and *ND1-cytb*. Significant negative values found in the total *An. gambiae* sample likely were due to inappropriate pooling of locations in Senegal and Kenya (see below), each with a high proportion of singletons. Interpretation of the significant negative value of Fu and Li's D from *An. gambiae* in Senegal will require sequence data from multiple unlinked loci (Simonsen *et al.* 1995). In the absence of compelling evidence to the contrary, we assume that variation detected in the mtDNA of both species is neutral and that selection is not responsible for the maintenance of shared polymorphisms between these species.

Geographic structure and gene flow within species: The locations sampled in Senegal and Kenya were separated by ~6000 km and were, respectively, ~7000 and ~2000 km removed from Kruger National Park, South Africa. Because they represent opposite extremes of the species' distributions on the African continent, it was expected that comparison of sequences sampled from these locations would show the maximum levels of genetic divergence.

The effect of distance on levels of gene flow was estimated from pairwise F_{ST} values. For both species and all comparisons among paired locations within 700 km of each other, these values were either zero or insignificantly small, as judged by the results of 500 random permutations. For *An. arabiensis*, even the F_{ST} values for South Africa *vs.* Kenya or Senegal were not significant. Only for Senegal *vs.* Kenya was significant genetic differentiation detected, with $F_{ST} = 0.018$ ($P < 0.02$) for *An. arabiensis*, and $F_{ST} = 0.085$ ($P < 0.002$) for *An. gambiae*. While significant, these values nevertheless corre-

TABLE 2

Summary statistics for mtDNA polymorphism

	n	S	h	π	θ_S	θ_U	Tajima's D	F and L's D
An. gambiae								
Total (*ND5*)	65	28	0.948	0.0038 (0.0021)	0.0089 (0.0028)	0.0181 (0.0039)	−1.836*	−2.46*
Senegal (*ND5*)	28	17	0.886	0.0038 (0.0021)	0.0066 (0.0025)	0.0107 (0.0032)	-1.467^{NS}	−2.97**
Kenya (*ND5*)	37	15	0.938	0.0035 (0.0019)	0.0054 (0.0021)	0.0093 (0.0027)	-1.158^{NS}	-0.28^{NS}
(*ND1-cytb*)	37	11	0.856	0.0031 (0.0019)	0.0050 (0.0020)	ND	-1.139^{NS}	-0.79^{NS}
An. arabiensis								
Total (*ND5*)	55	25	0.937	0.0046 (0.0024)	0.0082 (0.0027)	0.0202 (0.0044)	$-.1413^{NS}$	-1.48^{NS}
Senegal (*ND5*)	26	23	0.988	0.0051 (0.0027)	0.0091 (0.0034)	0.0209 (0.0056)	-1.602^{NS}	-1.84^{NS}
S. Africa (*ND5*)	5	3	0.70	0.0023 (0.0016)	0.0022 (0.0015)	0.0026 (0.0017)	0.262^{NS}	0.76^{NS}
Kenya (*ND5*)	23	8	0.806	0.0040 (0.0022)	0.0033 (0.0015)	0.0060 (0.0022)	0.684^{NS}	1.42^{NS}
(*ND1-cytb*)	23	3	0.708	0.0026 (0.0016)	0.0015 (0.0010)	ND	1.690^{NS}	0.99^{NS}

n, the number of sequences; S, the number of polymorphic sites; h, haplotype diversity; π, θ_S, θ_U, Tajima's D, F and L's D, are as defined in MATERIALS AND METHODS; ND, not computed. The total *An. arabiensis* sample included a single specimen from the Mananga colony as well as all field specimens from Malahlapanga, South Africa. Values in parentheses are SE. * $P < 0.05$; ** $P < 0.02$.

sponded to average levels of migration (N_m) in excess of two per generation, above the threshold required for genetic differentiation by genetic drift (SLATKIN, 1987).

Evidence for isolation by distance was examined in both species for each pair of locations up to 100 km apart, and for *An. arabiensis* over the entire continent, after pooling the six samples from western Kenya and the two from southwestern Senegal. Because N_m is undefined when $F_{ST} = 0$ (as it often was), pairwise F_{ST} values rather than N_m values were plotted against distance (Figure 3), and against log(distance) (not shown). Although the highest F_{ST} values were associated with comparisons across 2000 km, it was also true that some pairs of samples separated by 6000–7000 km showed $F_{ST} = 0$. No simple nor convincing pattern of isolation by distance was evident from these mtDNA sequences; variation in levels of differentiation was independent of distance. However, it should be noted that the coarse geographical sampling scheme, and small sample sizes at some locations (*e.g.*, coastal Kenya), may have limited the ability to detect isolation by distance.

Analyses based on this single, maternally inherited locus seemed to suggest that neither large geographic distances nor species designations play an important role in stratifying variation in *An. gambiae* and *An. arabiensis*. Evidence from allozyme and microsatellite markers supports this view. Allozyme frequency data from both species, collected from numerous locations throughout Africa (MILES, 1978), were analyzed using BIOSYS. F_{ST} analyses were carried out on five enzyme loci from *An. gambiae* and four from *An. arabiensis* sampled from West Africa (Senegal and The Gambia), from East Africa (Kenya), and from an *An. arabiensis* population from South Africa. These enzyme loci are autosomal; the esterases and Pgm2 have been mapped to chromosome *2* and Odh is linked to chromosome *3* (HUNT, 1987; CREWS-OYEN *et al.* 1993). In addition, microsatellite data from five loci representing all three chromosomes are available from *An. gambiae* populations from Senegal and Kenya (LEHMANN *et al.* 1996, 1997), and from 25 loci from both species in Kenya (KAMAU *et al.* 1997). Table 3 compares the F_{ST} estimates based on mtDNA to those derived from allozyme and microsatellite markers. Because sampling times and locations were different for the different markers and the calculation of F_{ST} or R_{ST} differed by marker, exact numbers cannot be appropriately compared. However, the concordant patterns of genetic differentiation indicated by the different markers strengthen the view that large geographic distances have not imposed a barrier to gene flow within species and emphasize the low level of divergence between species.

Pairwise difference distributions: If the frequency of mtDNA haplotype pairs that differ by i nucleotide sites, where $i \geq 0$, is plotted, the shape of the distribution can provide information about the history of changes in population size (SLATKIN and HUDSON, 1991; ROGERS and HARPENDING, 1992). Simulation studies suggest that for a stable population, the shape is expected to be

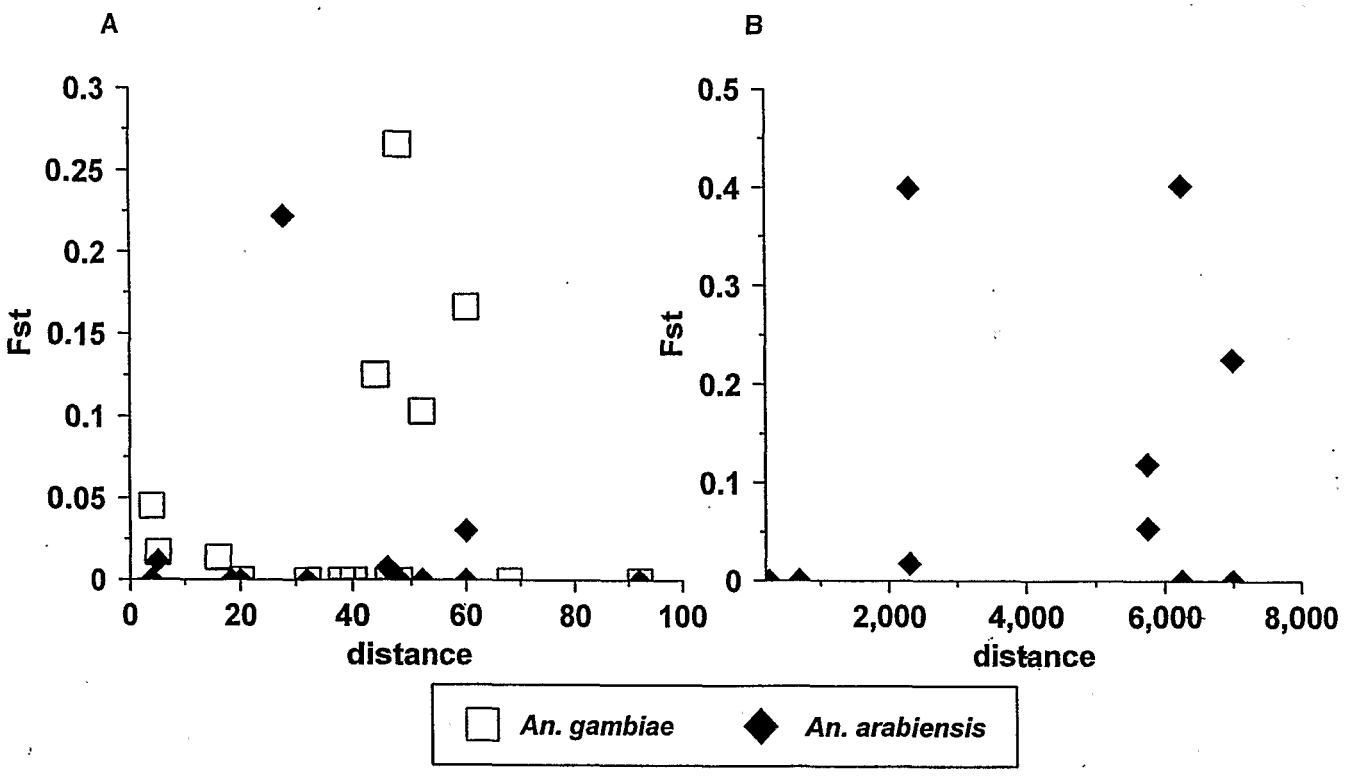

FIGURE 3.—Values of F_{st} plotted against distance (km) on a microgeographic scale for both species (A), or a macrogeographic scale for *An. arabiensis* (B). In each case, the F_{st} values were weighted by the harmonic mean of the sample sizes.

TABLE 3

Geographic structure and gene flow indicated by mtDNA (ND5), allozyme, and microsatellite markers

	mtDNA[a] F_{ST}	Allozymes F_{ST}	Microsatellites R_{ST}
An. gambiae			
5–100	0.017 (28.9)	0.006 (41.4)	0.013[c] (38.0)
250–700 km	0.032 (15.1)	0.025 (9.8)	0.037[d] (9.8)
2000–7000 km	0.085 (5.4)	0.030 (8.1)	0.036[e] (3.4)
An. arabiensis			
5–100 km	0 (⩾1)	—ND	—ND
250–700 km	0 (⩾1)	0.021 (11.7)	—ND
2000–7000 km	0.044 (10.8)	0.038 (6.3)	—ND
An. gambiae vs.			
An. arabiensis	0.093	0.072[b]	0.153[f]

Values in parentheses are *Nm*. ND, no data available.

[a] Summary F_{ST} values for each geographic range were an average of pairwise F_{ST} values, with each pairwise value weighted by the harmonic mean of sample sizes. In computing this average, negative pairwise F_{ST} values were not considered as zero, to better account for random noise. However, any negative summary values obtained after averaging were considered as zero.

[b] Average across four loci, excluding *ODH*. Including *ODH*, $F_{ST} = 0.191$.

[c] Average across five loci from LEHMANN *et al.* (1997).

[d] Average across nine loci (T. LEHMANN, unpublished data).

[e] Average across five loci (from LEHMANN *et al.* 1996).

[f] Median R_{ST} value for 25 loci (from KAMAU *et al.* 1997).

irregular because of stochastic lineage loss. By contrast, recent population expansion should produce a smooth unimodal distribution that approximates Poisson. For both *An. gambiae* and *An. arabiensis*, the distribution of pairwise differences for *ND5* sequences from Senegal and Kenya was plotted (Figure 4). Although all distributions except Kenyan *An. arabiensis* appeared unimodal, each strongly deviated from POISSON ($P \ll 0.01$). This result is consistent with population stationarity but does not necessarily rule out more complex models of population expansion (see ELLER and HARPENDING, 1996).

mtDNA phylogeography: With only 25 polymorphic sites that were informative for parsimony and more than twice that number of haplotypes, maximum parsimony analyses could not be usefully applied. Instead, a parsimony network was constructed in which the quantity minimized was not global tree length, but individual mutational connections between haplotypes (Figure 5). This approach better reflected the expectation that, for a collection of very closely related sequences, common ancestral sequences will be extant in the collection (CRANDALL *et al.* 1994). The probability of a parsimonious connection between haplotypes was supported across the network at $P \geq 0.97$ (equation 8 of TEMPLETON *et al.* 1992). However, the proportion of ambiguous connections indicates that at least some mutations are homoplasious. Indeed, although the network-based mutational distances between pairs of haplotypes were often in good agreement with observed distances, this was not always the case. For example, haplotypes 24

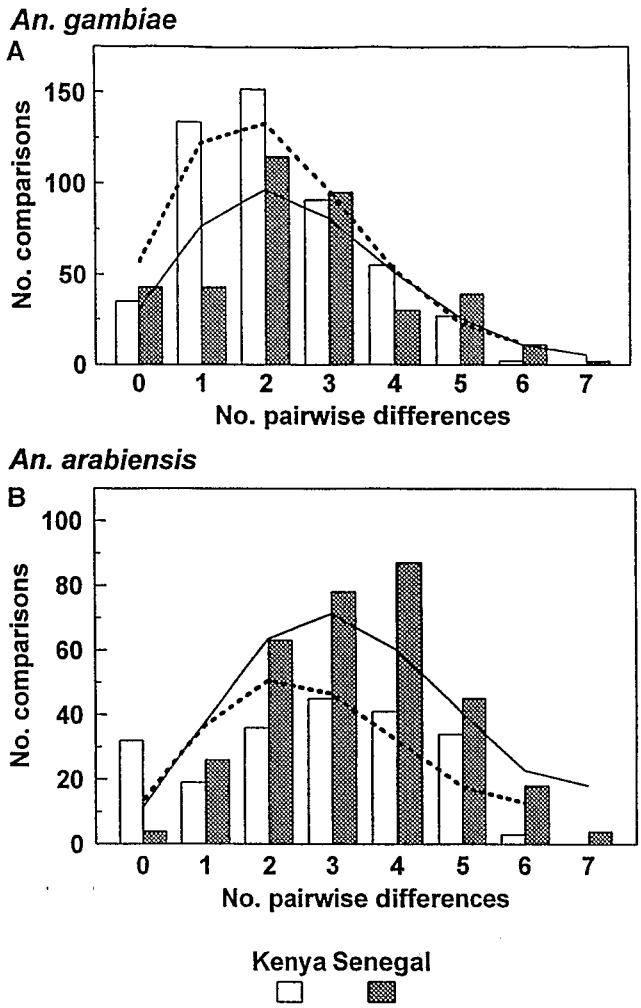

FIGURE 4.—Pairwise-difference distributions from mtDNA *ND5* sequences of *An. gambiae* (A) and *An. arabiensis* (B). Empirical distributions are shown as hatched bars (Senegal) or white bars (Kenya); their fit to a Poisson distribution is given by solid (Senegal) or dashed (Kenya) lines.

and 25 are seven steps apart in the network but only five steps apart when the sequences are compared directly.

In Figure 5, those haplotypes that were most frequent in the collection were generally positioned at interior nodes, from which many connections sprouted. In A, where haplotypes are color-coded according to species, it can be seen that most of the more frequent, ancestral haplotypes were shared between species, whereas terminal haplotypes (those at the tips) were rarely shared. So far, this pattern is most easily interpreted as sorting of shared ancestral polymorphism. However, assuming that terminal haplotypes were created by the most recent mutations, it is worth noting that three such haplotypes from *An. arabiensis* (8, 17, and 26) were connected most parsimoniously to haplotypes only found in *An. gambiae.* In B, the same network of haplotypes was coded according to geographic origin (Senegal, Kenya, or South Africa). This revealed a similar pattern, in which the more frequent ancestral haplotypes were generally shared between geographic regions and the terminal ones were not. However, there were three terminal haplotypes from Senegal (8, 17, and 19) that were connected to haplotypes found only in Kenya, even when all ambiguous connections were considered. It is striking that haplotypes 8 and 17 were also from *An. arabiensis* but connected to *An. gambiae* haplotypes.

DISCUSSION

Intraspecific mtDNA variation: Previously published studies of mtDNA variation within anopheline species have used restriction mapping of the entire molecule (*e.g.*, CONN *et al.* 1993, 1997; FREITAS-SIBAJEV *et al.* 1995; PERERA *et al.* 1995). With this technique, estimated levels of nucleotide diversity found for six species of nearctic and neotropical anopheline mosquitoes ranged from 0.0018 to 0.0085 (CONN *et al.* 1997). Nucleotide diversity estimates obtained by sequencing the ND5 gene fragment of *An. gambiae* and *An. arabiensis* (0.0038 and 0.0046, respectively) fall within this range. However, the pattern of mtDNA variation found in each geographic location in the *An. gambiae* complex is at least superficially distinct from that described for the other anophelines, in that no haplotype predominates and the majority of haplotypes are singletons. Furthermore, cooccurring haplotypes commonly differ at four to five nucleotide positions. This is a pattern consistent with relatively large continuous populations that do not experience severe dry season bottlenecks, a conclusion reached by TAYLOR *et al.* (1993) based on frequencies of chromosomal inversions in *An. arabiensis.* In the other anophelines examined to date, one or two haplotypes predominate at each geographic location, with rare ones differing by a single restriction site (CONN *et al.* 1993, 1997; FREITAS-SIBAJEV *et al.* 1995; PERERA *et al.* 1995). A possible explanation for this distinction is the higher resolution of sequencing *vs.* restriction mapping. It remains to be seen whether underlying differences in population size and structure also contribute.

Despite high levels of allelic diversity, we detected little to no geographic structure to the mtDNA variation within *An. gambiae* and *An. arabiensis,* even at distances of 6000 km. Although there are few obvious environmental barriers to gene flow, these results were unanticipated. In part, this was because the population genetics of these species has been studied mainly from the point of view of chromosomal inversion polymorphisms (*e.g.*, COLUZZI *et al.* 1985; TOURÉ *et al.* 1994), which are unlikely to be selectively neutral. For *An. gambiae,* in which the frequencies of alternative chromosomal arrangements in West Africa show significant temporal and spatial heterogeneities strongly correlated with environmental factors (COLUZZI *et al.* 1985; TOURÉ *et al.* 1994), the impression was highly structured populations. However, if natural selection were responsible for this pattern, it may "say more about the environmental conditions than about the gene flow regime of the species"

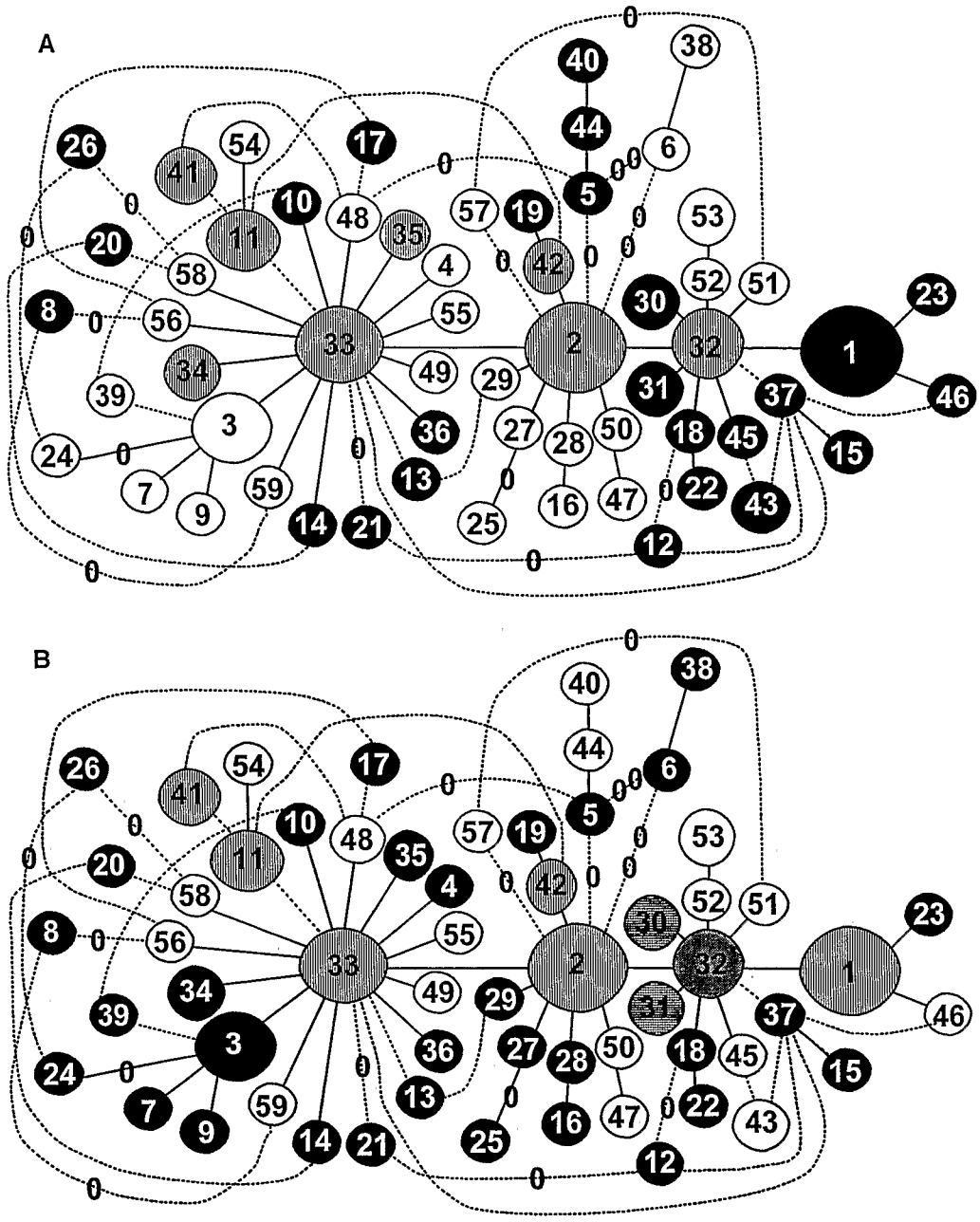

FIGURE 5.—Parsimony network for 59 mtDNA *ND5* haplotypes from *An. gambiae* and *An. arabiensis*. Numbered haplotypes are enclosed in circles roughly proportional to their frequency; "0" are missing intermediates. Each connection represents one nucleotide difference. Dashed lines indicate ambiguous connections. In A, haplotypes are color-coded according to taxonomy, with white circles and black circles representing haplotypes unique to *An. gambiae* or *An. arabiensis*, respectively. Hatched circles represent haplotypes common to both species. In B, haplotypes are color-coded according to geography, with white circles, black circles, and checked circles (haplotypes 30–31) representing haplotypes unique to Kenya, Senegal, and South Africa, respectively. Hatched circles (1–2, 11, 33, 41–42) represent haplotypes common to Senegal and Kenya; cross-hatched circle (32) represents a haplotype common to South Africa and Kenya.

(AVISE, 1994). Indeed, not only do patterns of allozyme and microsatellite variation concur with mtDNA, but estimates of gene flow based on these different classes of markers are remarkably similar in spite of mutation rates that may differ by several orders of magnitude (LEHMANN *et al.* 1996, 1997; Table 3).

One possible interpretation is that gene flow within both species is extensive enough to prevent differentiation across the African continent. An alternative explanation is that this pattern does not reflect current population structure, but the history of a recent range expansion across Africa by both species. The shape of the mismatch distributions (Figure 4) does not support rapid population growth from a small founder population, but is still compatible with range expansion by large and relatively stable populations, a scenario recently proposed for a similar pattern found in the Western European house mouse (NACHMAN *et al.* 1994). The parsimony network (Figure 5) fulfills a prediction of recent colonization in that the older, more frequent interior haplotypes are the most widespread geographically, whereas the newer terminal ones are geographically limited (CASTELLOE and TEMPLETON, 1994).

Is the dispersal ability of these mosquitoes consistent with that required to make gene flow a plausible explanation for low differentiation across Africa? One mark-release-recapture study measured the mean flight range at <1.6 km, with over 90% recaptured within 3 km of the release point (GILLIES, 1961). However, flight capacity of *An. gambiae* is as high as 7 km (HOLSTEIN, 1954), and a recent mark-release study gave considerably higher estimates of 350–650 meters per day per

adult mosquito, whose daily probability of survival was 0.8 (CONSTANTINI *et al.* 1996). NEIGEL and AVISE (1993) have modeled the expected geographic ranges for mtDNA lineages dispersed by a multigeneration "random walk." Assuming a per generation dispersal distance of 2–10 km, 12 generations per year, and mtDNA lineages 2–12 million generations old (using a molecular clock of 1.1–1.2% per million years per lineage; BROWER, 1994), the expected geographic range for these lineages actually exceeds 6000 km (Figure 4 of NEIGEL and AVISE, 1993), so that it is at least theoretically possible that equilibrium exists between gene flow and genetic drift across Africa. Beyond the possibility of active dispersion, it should also be noted that passive transport of *An. gambiae* over long distances by plane, boat, and truck has been recorded repeatedly both within and outside its normal distribution range (HOLSTEIN, 1954). When such transport leads to previously unoccupied territory with potential breeding sites, the process of colonization can be remarkably rapid. *An. gambiae* invaded Brazil in late 1929 or early 1930 and within 10 years had spread into an area with a ~230 km radius before it was eradicated (SOPER and WILSON, 1943).

Although active and passive dispersal may be sufficient explanations for the distribution of mtDNA variation in both species, their behavior makes it difficult to discount recent range expansion as an additional factor. *An. gambiae* and to a lesser degree *An. arabiensis* are "domestic" species, not only depending on humans for blood, but also for resting and breeding sites, since they rest in human habitations or storage sheds and breed in bare sunlit pools created by human activity such as footprints, animal hoofprints, rice fields, and irrigation ditches. The anthropophily of *An. gambiae* led COLUZZI *et al.* (1985, p 46) to observe that "it would be difficult to hypothesize its evolution and its wide diffusion in Tropical Africa in the absence of man." On behavioral grounds, it is tempting to speculate that *An. gambiae* is a relatively young species, whose origin may have coincided with the thousand-fold increase of human populations in Africa following the arrival of agriculture into West Africa within the last 9000 years (CAVALLI-SFORZA *et al.* 1993).

Interspecific mtDNA variation: Of the several dozen species of anophelines implicated in human malaria transmission, most are thought to belong to sibling species complexes (COLLINS and PASKEWITZ, 1995); thus the ability to discriminate members of a complex has a practical dimension that has motivated many studies. Where restriction fragment length polymorphism surveys of mtDNA variation in other anopheline species complexes have been made, in *An. dirus* (YASOTHORNSRIKUL *et al.* 1988), *An. freeborni* (COLLINS *et al.* 1990), *An. quadrimaculatus* (MITCHELL *et al.* 1992) and *An. albitarsis* (NARANG *et al.* 1993), fixed differences distinguished member species in all cases except the *An. freeborni* complex (COLLINS *et al.* 1990). *An. gambiae* and *An. arabiensis* constitute another exception, since there were no fixed nucleotide differences and multiple shared haplotypes between them.

How likely is it that these taxa share identical mtDNA haplotypes solely on the basis of retained ancestral polymorphism? The expectation that fixed mtDNA differences will distinguish closely related species depends on the time since species splitting and female effective population size. The shorter the time and the larger the population size, the greater the likelihood of mtDNA polyphyly or paraphyly between taxa (AVISE, 1994). Assuming that all gene flow between species permanently ceased at the time of species splitting, the probability that both taxa retain identical mtDNA haplotypes from the ancestral population is very slight beyond $4n$ generations, where n is carrying capacity or female population size (AVISE *et al.* 1984). The long-term effective population size of *An. gambiae* has been estimated from microsatellite frequencies across nine loci to be ~240,000, using a conservative average mutation rate of 10^{-5} (T. LEHMANN, unpublished data). Using this estimate and assuming stable-sized populations and selective neutrality, it is very unlikely for ancestral mtDNA haplotypes to have persisted in either taxa beyond 480,000 generations (~40,000 years). This is much less than the mean time to coalescence for the most divergent haplotypes in either species: ~510,000 years with a standard deviation of ~180,000 years (calculated after TEMPLETON, 1993, using a mutation rate of 10^{-8} per year). Based on a gamma distribution, the 95% confidence limits about these means, 219,000–914,000 years, do not include 40,000 years. However, there is no external calibration for the mtDNA mutation rate in *An. gambiae,* and if a 10-fold higher mutation rate is used, the mean time to coalescence is ~51,000 years with a standard deviation of 18,000 years, and the 95% confidence limits of 21,900–91,400 years easily include 40,000 years. With conservative estimates of mutation rates, this coalescent approach does not rule out shared ancestral polymorphism as the explanation for shared haplotypes. However, if retained ancestral polymorphism were accepted as a sufficient explanation, nuclear gene polymorphisms should be shared even more extensively than those of mtDNA (because of their larger effective population sizes), and this has not yet been found outside of shared inversions (BESANSKY *et al.* 1994; MATHIOPOULOS *et al.* 1995; GARCIA *et al.* 1996). Therefore, we believe that transfer of mtDNA across species boundaries provides a plausible explanation for haplotypes shared between these taxa.

Frequent examples of natural hybridization and introgression exist in both plants and animals (HARRISON, 1993; AVISE, 1994; ARNOLD, 1997), particularly among closely related taxa. What is unusual in the case of *An. gambiae* and *An. arabiensis* is the apparent absence of any hybrid zones and the geographic scale of mtDNA

introgression (but see MASON *et al.* 1995), such that without independent molecular or cytogenetic markers, these isomorphic taxa would not be distinguishable. Although potentially fertile female hybrids have been found at rates of one to two per thousand females examined (COLUZZI *et al.* 1979), rates theoretically high enough to homogenize interspecific variation, their relative fitness in natural populations has not been studied. Our data indicate that at least some proportion of hybrids must successfully backcross to either parent species, and depending upon the direction of the backcross, transfer mtDNA across species boundaries. This raises the possibility that portions of the nuclear genome could be transferred as well. Thus, both ancestral polymorphism and more recent introgressive hybridization challenge the interpretation of gene flow within and between *An. gambiae* and *An. arabiensis*. The rewards of this challenge lie not only in information about the nature of divergence at the population-species interface, but in information critical to the successful application and evaluation of future genetic control strategies that target these dangerous vectors.

We are grateful to A. CORNEL for providing specimens of the *An. arabiensis* Mananga colony from the South African Institute for Medical Research and to A. HIGHTOWER for SAS support. KEITH CRANDALL kindly calculated the probability of parsimonious haplotype connections for the estimated cladograms. We thank BRIAN HOLLOWAY and staff of the NCID Biotechnology Core Facility for synthesizing the oligonucleotide primers. T.L. was supported by an American Society for Microbiology Fellowship. This work was supported by a grant from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases to N.J.B. and a grant from the JOHN D. and CATHERINE T. MACARTHUR Foundation to F.H.C.

LITERATURE CITED

ARNOLD, M. L., 1997 *Natural Hybridization and Evolution.* Oxford University Press, New York.

AVISE, J. C., 1994 *Molecular Markers, Natural History and Evolution.* Chapman and Hall, New York.

AVISE, J. C., J. E. NEIGEL and J. ARNOLD, 1984 Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. **20:** 99–105.

BEARD, C. B., D. MILLS HAMM and F. H. COLLINS, 1993 The mitochondrial genome of the mosquito *Anopheles gambiae:* DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol. Biol. **2:** 103–124.

BESANSKY, N. J., J. R. POWELL, A. CACCONE, D. M. HAMM, J. A. SCOTT *et al.*, 1994 Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. Proc. Natl. Acad. Sci. USA **91:** 6885–6888.

BRAACK, L. E. O., M. COETZEE, R. H. HUNT, H. BIGGS, A. CORNEL *et al.*, 1994 Biting pattern and host-seeking behavior of *Anopheles arabiensis* (Diptera: Culicidae) in northeastern South Africa. J. Med. Entomol. **31:** 333–339.

BROWER, A. V. Z., 1994 Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. USA **91:** 6491–6495.

CACCONE, A., B. A. GARCIA and J. R. POWELL, 1996 Evolution of the mitochondrial DNA control region in the *Anopheles gambiae* complex. Insect Mol. Biol. **5:** 51–59.

CASTELLOE, J., and A. R. TEMPLETON, 1994 Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. **3:** 102–113.

CAVALLI-SFORZA, L. L., P. MENOZZI and A. PIAZZA, 1993 Demic expansions and human evolution. Science **259:** 639–646.

COETZEE, M., R. H. HUNT, L. E. O. BRAACK and G. DAVIDSON, 1993 Distribution of mosquitoes belonging to the *Anopheles gambiae* complex, including malaria vectors, south of latitude 15°S. S. Afr. J. Sci. **89:** 227–231.

COLLINS, F. H., and S. M. PASKEWITZ, 1995 Malaria: current and future prospects for control. Annu. Rev. Entomol. **40:** 195–219.

COLLINS, F. H., A. MENDEZ, M. O. RASSMUSSEN, P. C. MEHAFFEY, N. J. BESANSKY *et al.*, 1987 A ribosomal RNA probe differentiates member species of the *Anopheles gambiae* complex. Am. J. Trop. Med. Hyg. **37:** 37–41.

COLLINS, F. H., C. H. PORTER and S. E. COPE, 1990 Comparison of rDNA and mtDNA in the sibling species *Anopheles freeborni* and *A. hermsi*. Am. J. Trop. Med. Hyg. **42:** 417–423.

COLUZZI, M., A. SABATINI, V. PETRARCA and M. A. DI DECO, 1979 Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. Trans. R. Soc. Trop. Med. Hyg. **73:** 483–487.

COLUZZI, M., V. PETRARCA and M. A. DI DECO, 1985 Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae.* Bull. Zool. **52:** 45–63.

CONN, J., A. F. COCKBURN and S. E. MITCHELL, 1993 Population differentiation of the malaria vector *Anopheles aquasalis* using mitochondrial DNA. J. Hered. **84:** 248–253.

CONN, J., S. E. MITCHELL and A. F. COCKBURN, 1997 Mitochondrial DNA variation within and between two species of neotropical anopheline mosquiotes (Diptera: Culicidae). J. Hered. **88:** 98–107.

CONSTANTINI, C., S.-G. LI, A. DELLA TORRE, N. SAGNON, M. COLUZZI *et al.*, 1996 Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a West African Sudan savanna village. Med. Vet. Entomol. **10:** 203–219.

CRANDALL, K. A., A. R. TEMPLETON and C. F. SING, 1994 Intraspecific phylogenetics: problems and solutions, pp. 273–297 in *Models in Phylogeny Reconstruction,* edited by R. W. SCOTLAND, D. J. SIEBERT and D. M. WILLIAMS. Clarendon Press, Oxford.

CREWS-OYEN, A. E., V. KUMAR and F. H. COLLINS, 1993 Association of two esterase genes, a chromosomal inversion, and susceptibility to *Plasmodium cynomolgi* in the African malaria vector *Anopheles gambiae.* Am. J. Trop. Med. Hyg. **49:** 341–347.

DELLA TORRE, A., L. MERZAGORA, J. R. POWELL and M. COLUZZI, 1997 Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. Genetics **146:** 239–244.

ELLER, E., and H. C. HARPENDING, 1996 Simulations show that neither population expansion nor population stationarity in a West African population can be rejected. Mol. Biol. Evol. **13:** 1155–1157.

FONTENILLE, D., L. LOCHOUARN, N. DIAGNE, C. SOKHNA, J. J. LEMASSON *et al.*, 1997a High annual and seasonal variations in malaria transmission by anopheline and vector species composition in Dielmo, a holoendemic area in Senegal. Am. J. Trop. Med. Hyg. **56:** 247–253.

FONTENILLE, D., L. LOCHOUARN, M. DIATTA, C. SOKHNA, I. DIA *et al.*, 1997b A 4-year entomological study of seasonal malaria in Ndiop, a mesoendemic area in Senegal: large annual variations of transmission, close bionomics of *Anopheles gambiae* and *An. arabiensis.* Trans. Royal Soc. Trop. Med. Hyg. (in press).

FREITAS-SIBAJEV, M. G. R., J. CONN, S. E. MITCHELL, A. F. COCKBURN, J. A. SEAWRIGHT *et al.*, 1995 Mitochondrial DNA and morphological analyses of *Anopheles darlingi* (Diptera: Culicidae) populations from Brazil. Mosq. Syst. **27:** 78–99.

FU, Y.-X., 1994 A phylogenetic estimator of effective population size or mutation rate. Genetics **136:** 685–692.

FU, Y.-X., and W.-H. LI, 1994 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

GARCÍA, B. A., A. CACCONE, K. D. MATHIOPOULOS and J. R. POWELL, 1996 Inversion monophyly in African anopheline malaria vectors. Genetics **143:** 1313–1320.

GILLIES, M. T., 1961 Studies on the dispersion and survival of *Anopheles gambiae* Giles in East Africa, by means of marking and release experiments. Bull. Entomol. Res. **52:** 99–127.

GILLIES, M. T., and B. DE MEILLON, 1968 *The Anophelinae of Africa south of the Sahara.* Publications of the South African Institute for Medical Research No. 54, Johannesburg.

HARRISON, R. G., 1993 *Hybrid Zones and the Evolutionary Process.* Oxford University Press, New York.

HOLSTEIN, M. H., 1954 *Biology of Anopheles gambiae.* World Health Organization Monograph Series No. 9, Geneva.

HUDSON, R. R., M. SLATKIN and W. P. MADDISON, 1992a Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

HUDSON, R. R., D. D. BOOS and N. L. KAPLAN, 1992b A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

HUNT, R. H., 1987 Location of genes on chromosome arms in the *Anopheles gambiae* group of species and their correlation to linkage data for other anopheline mosquitoes. Med. Vet. Entomol. **1:** 81–88.

KAMAU, L., W. A. HAWLEY, T. LEHMANN, A. S. S. ORAGO, A. CORNEL *et al.*, 1997 Use of short tandem repeats for analysis of genetic variability in sympatric populations of *Anopheles gambiae* and *Anopheles arabiensis.* Heredity (in press).

KUMAR, S., K. TAMURA and M. NEI, 1993 MEGA: *Molecular Evolutionary Genetics Analysis, version 1.01.* The Pennsylvania State University, University Park.

LEHMANN, T., W. A. HAWLEY, L. KAMAU, D. FONTENILLE, F. SIMARD *et al.*, 1996 Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. Heredity **77:** 192–208.

LEHMANN, T., N. J. BESANSKY, W. A. HAWLEY, T. G. FAHEY, L. KAMAU *et al.*, 1997 Microgeographic structure of Anopheles gambiae in western Kenya based on mtDNA and microsatellite loci. Mol. Ecol. **6:** 243–253.

LEMASSON, J. J., D. FONTENILLE, L. LOCHOUARN, I. DIA, F. SIMARD *et al.*, 1997 Comparison of behaviour and vector efficiency of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae) in Barkedji, a Sahelian area of Senegal. J. Med. Entomol. **34:** 396–403.

MASON, D. J., R. K. BUTLIN and P. GACESA, 1995 An unusual mitochondrial DNA polymorphism in the *Chorthippus biguttulus* species group (Orthoptera: Acrididae). Mol. Ecol. **4:** 121–126.

MATHIOPOULOS, K. D., J. R. POWELL and T. F. MCCUTCHAN, 1995 An anchored restriction-mapping approach applied to the genetic analysis of the *Anopheles gambiae* malaria vector complex. Mol. Biol. Evol. **12:** 103–112.

MCLAIN, D. K., F. H. COLLINS, A. DAVID BRANDLING-BENNETT and J. B. O. WERE, 1989 Microgeographic variation in rDNA intergenic spacers of *Anopheles gambiae* in western Kenya. Heredity **62:** 257–264.

MILES, S. J., 1978 Enzyme variation in the Anopheles gambiae Giles group of species (Diptera: Culicidae). Bull. Ent. Res. **68:** 85–96.

MITCHELL, S. E., S. K. NARANG, A. F. COCKBURN, J. A. SEAWRIGHT and M. GOLDENTHAL, 1992 Mitochondrial and ribosomal DNA variation among members of the *Anopheles quadrimaculatus* (Diptera: Culicidae) species complex. Genome **35:** 939–950.

NACHMAN, M. W., S. N. BOYER, J. B. SEARLE and C. F. AQUADRO, 1994 Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus.* Genetics **136:** 1105–1120.

NARANG, S. K., T. A. KLEIN, O. P. PERERA, J. B. LIMA and A. T. TANG, 1993 Genetic evidence for the existence of cryptic species in the *Anopheles albitarsus* complex in Brazil: allozymes and mitochondrial DNA restriction fragment length polymorphisms. Biochem. Genet. **31:** 97–112.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NEIGEL, J. E., and J. C. AVISE, 1993 Application of a random walk model to geographic distributions of animal mitochondrial DNA variation. Genetics **135:** 1209–1220.

PAPE, T., 1992 Cladistic analyses of mosquito chromosome data in *Anopheles* subgenus *Cellia* (Diptera: Culicidae). Mosquito Systematics **24:** 1–11.

PERERA, O. P., S. E. MITCHELL, A. F. COCKBURN and J. A. SEAWRIGHT, 1995 Variation in mitochondrial and ribosomal DNA of *Anopheles quadrimaculatus* species A (Diptera: Culicidae) across a wide geographic range. Ann. Entomol. Soc. Am. **88:** 836–845.

ROGERS, A. R., and H. HARPENDING, 1992 Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9:** 552–569.

ROZAS, J., and R. ROZAS, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Applic. Biosci. **13:** 307–311.

SAS INSTITUTE INC., 1990 *SAS Language: References.* Version 6, 1st edition. Cary, NC.

SCOTT, J. A., W. G. BROGDON and F. H. COLLINS, 1993 Identification of single specimens of the Anopheles gambiae complex by the polymerase chain reaction. Am. J. Trop. Med. Hyg. **49:** 520–529.

SIMONSEN, K. L., G. A. CHURCHILL and C. F. AQUADRO, 1995 Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141:** 413–429.

SLATKIN, M., 1987 Gene flow and the geographical structure of populations. Science **236:** 787–792.

SLATKIN, M., and R. R. HUDSON, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129:** 555–562.

SOPER, F. L., and D. B. WILSON, 1943 *Anopheles gambiae in Brazil 1930 to 1940.* The Rockefeller Foundation, New York.

SWOFFORD, D. L., and R. B. SELANDER, 1989 *BIOSYS-1. A Computer Program for the Analysis of Allelic Variation in Population Genetics and Biochemical Systematics.* Release 1.7. University of Illinois, Urbana.

TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

TAJIMA, F., 1993 Measurement of DNA polymorphism, pp. 37–59 in *Mechanisms of Molecular Evolution,* edited by N. TAKAHATA and A. G. CLARK. Sinauer Associates, Sunderland, MA.

TAYLOR, C. E., Y. T. TOURE, M. COLUZZI and V. PETRARCA, 1993 Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. Med. Vet. Entomol. **7:** 351–357.

TEMPLETON, A. R., 1993 The "Eve" hypothesis: a genetic critique and reanalysis. Am. Anthropol. **95:** 51–72.

TEMPLETON, A. R., K. A. CRANDALL and C. F. SING, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

TOURÉ, Y. T., V. PETRARCA, S. F. TRAORÉ, A. COULIBALY, H. M. MAIGA *et al.*, 1994 Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s.str. in Mali, West Africa. Genetica **94:** 213–223.

YASOTHRONSRIKUL, S., S. PANYIM and R. ROSENBERG, 1988 Diagnostic restriction fragment patterns of DNA from the isomorphic species of *Anopheles dirus.* Southeast Asian J. Trop. Med. Public Health **19:** 703–708.

Communicating editor: J. HEY